

Prediction of Ozone Concentration in Semi-Arid Areas of China Using a Novel Hybrid Model

W. Y. Zhang^{1,2}, J. J. Wang^{1,2*}, X. Liu^{1,2}, and J. Z. Wang³

¹Key Laboratory for Semi-Arid Climate Change of the Ministry of Education, College of Atmospheric Sciences, Lanzhou University, Lanzhou 730000, China

²Key Laboratory of Arid Climatic Change and Reducing Disaster of Gansu Province, Lanzhou 730000, China

³School of Mathematics and Statistics, Lanzhou University, Lanzhou 730000, China

Received 25 April 2012; revised 8 April 2013; accepted 5 September 2013; published online 25 September 2013

ABSTRACT. The semi-arid regions of China support delicately balanced ecosystems that are likely to be significantly affected by small changes in the level of atmospheric ozone. To predict ozone concentrations in such areas, this paper proposes a new hybrid forecasting model called an ARMT-CPSO-BP neural network, which is based on the association rule mining technique (ARMT), a chaotic particle swarm optimization algorithm (CPSO) and a back-propagation (BP) neural network (CPSO-BP neural network). This paper first uses the ARMT to find correlations between meteorological variables and ozone concentrations. The significant correlation coefficients are then fed into the CPSO-BP neural network to obtain a prediction. Finally, the predicted results are compared with the results obtained from the BP neural network and the regression model. The comparison shows that the proposed hybrid model is superior to both the BP neural network and the regression model. The hybrid model reduces the square root mean square error (RMSE) by 31.6% compared to the BP model and by 23.8% compared to the regression model. The hybrid model is promising for forecasting ozone concentration in arid and semi-arid areas.

Keywords: ozone prediction, association rule mining technique, neural network, chaotic, particle swarm optimization

1. Introduction

Ozone is a pollutant at ground level and has been found to be harmful to human beings. Ozone impacts not only humans but also the wellbeing of the flora and fauna of a region. Many researchers have also found that elevated ozone levels can reduce field-grown crop production (Clarke et al., 1990; Benton et al., 2000; Feng and Kobayashi, 2009). In recent years, the concentration of ozone has been greatly increasing (Meleux et al., 2007; Jaffe and Ray, 2007; Tanimoto, 2009). Thus, it has become increasingly important to devise control and public-warning strategies for ozone levels. Techniques that can predict the concentration of ozone are crucial to the development of these strategies.

Due to the inherent spatial and temporal variation in emissions, the influence of meteorological conditions, and uncertainties associated with initial and boundary conditions, it is difficult to use the theory of dynamical systems or chemical reaction models to predict ozone levels. Chen et al. (1998) constructed a model for the prediction of ozone levels by creating a

multidimensional phase space map from observed ozone concentrations based on the dynamical systems theory. Trevino and Mendez (1999) proposed another model that uses a set of simple chemical reactions for the NO_x and hydrocarbon photolysis to study the transient ozone generation process in polluted urban areas. Both of these models can be used to predict ozone levels.

The relationship between weather and ozone levels is highly complex and nonlinear. Neural networks have a strong ability to describe such problems, and thus, in recent years, more and more neural network methods have been applied to ozone forecasting. For example, Elkamel et al. (2001) presented an artificial neural network model that is able to predict ozone concentrations as a function of meteorological conditions and precursor concentrations. Lu et al. (2006) developed a two-stage neural network to predict ozone concentrations from meteorological conditions. Dutot et al. (2007) used a neural network combined with a neural classifier for real-time forecasting of maximum ozone levels in an urban atmosphere in the center of France. Prybutok et al. (2000) developed a neural network model for forecasting daily maximum ozone levels and compared the new model with two conventional statistical models, regression and autoregressive integrated moving average (ARIMA). The comparison showed that the neural network model outperformed both the regression and the ARIMA models.

* Corresponding author. Tel.: +86 931 8911031; fax: +86 931 8914278.
E-mail address: jujiewang@126.com (J. J. Wang).

Although neural networks can improve prediction accuracy (Wilson et al., 2002) to a certain extent, they suffer from many limitations (Abdel-Aal et al., 2009). One serious limitation is the difficulty in determining the optimum network topology and numerous design parameters, such as the number and size of the hidden layers, the type of neuron transfer functions for the various layers, the learning rate and momentum coefficient, and the training stop criteria to avoid over-fitting and ensure adequate generalization for new data. Generally speaking, there are too many choices to make in determining these parameters, with little guidance available, and designers often resort to trial and error, which can be tedious and time consuming.

Kennedy and Eberhart (1995) proposed a method for optimizing the structure of an artificial neural network to find the minimum error without having to adjust the parameters of the neural network. An improved particle swarm optimization (PSO) with chaotic local searching was adopted to help avoid local optima and to enhance learning capability. Tang et al. (2010) proposed a novel method for training a support vector machine (SVM) using chaos particle swarm optimization (CPSO). Their results show that their training method can meet the requirements of practical application.

Inspired by the above literature, this paper presents a hybrid method for ozone level forecasting, called an ARMT-CPSO-BP neural network, that combines the ARMT and a CPSO-BP neural network. With this method, the ARMT is used to reveal association rules between the meteorological variables and ozone concentrations. The CPSO-BP neural network improves the model's ability to determine the optimal nonlinear mapping between meteorological factors and ozone levels. The CPSO can also adjust the parameters of the neural network by itself. The Semi-Arid Climate and Environment Observatory of Lanzhou University (SACOL) (Huang et al., 2008) is on the Loess Plateau in Northwest China. We use the data provided by SACOL to study whether the proposed model (ARMT-CPSO-BP) can improve prediction accuracy in forecasting ozone concentrations.

The rest of this paper is organized as follows. In Section 2, the hybrid forecasting method and performance indices are proposed. A test case is used as an illustrative example for testing the models in Sections 3 and 4. Finally, Section 5 presents conclusions.

2. Hybrid Forecasting Method: the ARMT-CPSO-BP Neural Network

In this section, an overview of the ARMT-CPSO-BP method is presented. First, the association rule mining technique, BP neural network and chaotic particle swarm optimization algorithm are briefly introduced, and then the ARMT-CPSO-BP model is proposed. Finally, to quantitatively determine the best model, performance indices are proposed.

2.1. Association Rule Mining Technique

2.1.1. Basic Principles of the Association Rule Mining Technique

Let $I = \{I_1, I_2, \dots, I_m\}$ be a set of binary attributes, which are called items. Let T be a database of transactions. Each transaction t is represented as a binary vector, with $t[k] = 1$ if t bought the item I_k , and $t[k] = 0$ otherwise. There is one tuple in the database for each transaction. Let X be a subset of I . We say that a transaction t satisfies X if, for all items I_k in X , $t[k] = 1$ (Martínez-Ballesteros et al., 2011).

By an association rule, we mean an implication of the form $X \Rightarrow I_j$, where X is a set of some items in I , and I_j is a single item in I that is not present in X . The rule $X \Rightarrow I_j$ is satisfied in the set of transactions T with a confidence factor of $0 \leq c \leq 1$ if at least $c\%$ of transactions in T that satisfy X also satisfy I_j . We will use the notation $X \Rightarrow I_j / c$ to specify that the rule $X \Rightarrow I_j$ has a confidence factor of c (Martínez-Ballesteros et al., 2010).

Given the set of transactions T , we are interested in generating all the rules that satisfy the following two additional types of constraints:

(I) Syntactic Constraints: These constraints involve restrictions on items that can appear in a rule. For example, we may be interested only in rules that have a specific item I_x appearing in the consequent or rules that have a specific item I_x appearing in the antecedent. Combinations of the above constraints are also possible – we may request all the rules that have items from a predefined itemset X appearing in the consequent and items from another itemset Y appearing in the antecedent.

(II) Support Constraints: These constraints concern the number of transactions in T that support a rule. The support for a rule is defined to be the fraction of transactions in T that satisfy the union of items in the consequent and antecedent of the rule. Support should not be confused with confidence. While confidence is a measure of the rule's strength, support corresponds to statistical significance. In addition to statistical significance, another motivation for using support constraints is that we are usually interested only in rules with support above some minimum threshold for business reasons. If there is not enough support, this means that the rule is not worth consideration or that it is simply less important (see later for more detail).

In this formulation, the problem of rule mining can be decomposed into the following two subproblems:

(I) Generate all combinations of items that have fractional transaction support above a certain threshold called *minsupport*. Call those combinations large itemsets and all other combinations that do not meet the threshold small itemsets.

Syntactic constraints further constrain the admissible combinations. For example, if only rules involving an item I_x in the antecedent are of interest, then it is sufficient to generate only those combinations that contain I_x .

(II) For a given large itemset $Y = \{I_1, I_2, \dots, I_k, \dots, I_m\}$, $k \geq 2$, generate all the rules (with a maximum of k rules) that use items from the set I_1, I_2, \dots, I_k . The antecedent of each of these rules will be a subset X of Y such that X has $k - 1$ items, and the consequent will be the item $Y - X$. To generate a rule $X \Rightarrow I_j / c$ where $X = I_1, I_2, \dots, I_{j-1}, I_{j+1}, \dots, I_k$, take the support of Y and divide it by the support of X . If the ratio is greater than c , then

the rule is satisfied with a confidence factor of c ; otherwise the rule is not satisfied.

Note that if the itemset Y is large, then every subset of Y will also be large, and we must have their support counts available as a result of the solution of the first subproblem. Additionally, all rules derived from Y must satisfy the support constraint because Y satisfies the support constraint, and Y is the union of items in the consequent and antecedent of every such rule.

Having determined the large itemsets, the solution to the second subproblem is straightforward. In the next section, we focus on the first subproblem.

2.1.2. Discovering Large Itemsets

Figure 1 shows the template algorithm for finding large itemsets. Given a set of items I , an itemset $X + Y$ of items in I is said to be an extension of the itemset X if $X \cap Y = \phi$. The parameter dbsize is the total number of tuples in the database.

```

Procedure LargeItemsets
Begin
  let Large set  $L = \phi$ ;
  let Frontier set  $F = \{\phi\}$ ;
  while  $F \neq \phi$  do begin
    --make a pass over the database
    let Candidate set  $C = 0$ ;
    forall database tuples  $t$  do
      forall itemsets  $f$  in  $F$  do
        if  $t$  contains  $f$  then begin
          let  $C_f$  = candidate itemsets that are
            extensions of  $f$  and contained in  $t$ ;
          forall itemsets  $cf$  in  $C_f$  do
            if  $cf \in C$  then
               $C_f$ .count =  $C_f$ .count + 1;
            else begin
               $C_f$ .count = 0;
               $C = C + C_f$ ;
            end
          end
        end
      let  $F = \phi$ ;
      forall itemsets  $c$  in  $C$  do begin
        if  $\text{count}(c)/\text{dbsize} > \text{rminsupport}$  then
           $L = L + c$ ;
        if  $c$  should be used as a frontier then
           $F = F + c$ ;
      end
    end
  end

```

Figure 1. The template algorithm.

The algorithm makes multiple passes through the database. The frontier set for a pass consists of those itemsets that are extended during the pass. In each pass, the support for certain itemsets is measured. These itemsets, called candidate itemsets, are derived from the tuples in the database and the itemsets contained in the frontier set.

Associated with each itemset is a counter that stores the number of transactions in which the corresponding itemset has appeared. This counter is initialized at zero when an itemset is created.

Initially, the frontier set consists of only one element, which is the empty set. At the end of a pass, the support for a candidate itemset is compared with minsupport to determine if it is a large itemset. At the same time, it is determined if this itemset should be added to the frontier set for the next pass. The algorithm terminates when the frontier set becomes empty. The support count for the itemset is preserved when an itemset is added to the large/frontier set (Agrawal et al., 1993).

2.2. BP Neural Network

Artificial neural networks are data-driven, non-parametric models and can be useful for nonlinear identification and time series prediction. Back-propagation (BP) neural networks, which are the most commonly used neural networks, take inputs only from the previous layer and send outputs only to the next layer (Wang, 2009; Zhang and Lenan, 2009; Geethanjali et al., 2008). A three-layer BP neural network is shown in Figure. 2, where x_j ($j = 1, 2, \dots, n$) represents the input variables, z_i ($i = 1, 2, \dots, m$) represents the outputs of the neurons in the hidden layer and y_t ($t = 1, 2, \dots, l$) represents the outputs of the neural network. The back-propagation process determines the weights for the connections between the nodes based on data training, yielding a minimized least mean square error between the actual values and those estimated by the neural network's output. The connection weights are assigned random initial values. The error between the predicted and actual output values is back-propagated via the network to update the weights. The neural network must be trained before being applied for forecasting. We can infer a training process to update these weighted values as described by the following equations, which can be divided into two steps (Pérez and Reyes, 2006):

(I) Hidden layer stage: The outputs of all the neurons in the hidden layer are calculated by the following steps:

$$net_i = \sum_{j=0}^n w_{ij} x_j \quad i = 1, 2, \dots, m \quad (1)$$

$$z_i = f_H(net_i) \quad i = 1, 2, \dots, m \quad (2)$$

Here, net_i is the activation value of the i -th node, w_{ij} is the linkage weight between the i -th node and the j -th node, z_j is the output of the hidden layer, and f_H is the activation function of a node, usually a sigmoid function, which is described as follows:

$$f_H(x) = \frac{1}{1 + \exp(-x)} \quad (3)$$

(II) Output stage: The outputs of all the neurons in the output layer are given as follows:

$$y_t = f_t \left(\sum_{i=0}^m w_{it} z_i \right) \quad t = 1, 2, \dots, l \quad (4)$$

Here, $f_t (t = 1, 2, \dots, l)$ is the activation function, usually a linear function. All weights are initially assigned random values and are traditionally modified by the delta rule according to the training samples.

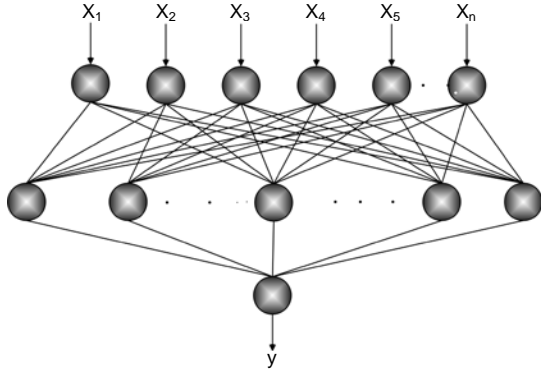


Figure 2. The structure of a BP neural network.

2.3. Chaotic Particle Swarm Optimization Algorithm

2.3.1. Particle Swarm Optimization Method

The particle swarm optimization (PSO) method was first proposed by Kennedy and Eberhart in 1995 (Kennedy and Eberhart, 1995) and has been widely used by many researchers (Cai et al., 2007). A PSO system starts with a random initialization of a population (swarm) in the problem search space, with each individual particle being a candidate solution to the problem. Each swarm particle flies through the space looking for the optimal landing position. During the search process, each particle accumulates experience by tracking and memorizing the best position it has encountered. The position of each particle in each generation is adjusted according to its velocity, which is adjusted dynamically based on the particle's own flying experience (its own best position) as well as the experience of neighboring particles (the best position of the entire swarm). Let the position, velocity, and best position of the i -th particle respectively be:

$$X_{(k)i} = [x_{(k)i,1}, x_{(k)i,2}, \dots, x_{(k)i,n}] \quad (5)$$

$$V_{(k)i} = [v_{(k)i,1}, v_{(k)i,2}, \dots, v_{(k)i,n}] \quad (6)$$

$$P_{(k)i} = [p_{(k)i,1}, p_{(k)i,2}, \dots, p_{(k)i,n}] \quad (7)$$

where $i = 1, 2, \dots, N$, k represents the parameters of the neural network for the PSO to determine, N is the number of parti-

cles, n is the dimension of the search space, and the global best position for all particles in the swarm is:

$$P_{(k)g} = [p_{(k)g,1}, p_{(k)g,2}, \dots, p_{(k)g,n}] \quad i = 1, 2, \dots, N \quad (8)$$

Then, the new velocity of the i -th particle can be computed by:

$$V_{(k)i}(t+1) = lV_{(k)i}(t) + q_1 \text{rand}(\bullet)(P_{(k)i} - X_{(k)i}(t)) + q_2 \text{Rand}(\bullet)(P_{(k)g} - X_{(k)i}(t)) \quad i = 1, 2, \dots, N \quad (9)$$

where l denotes the inertia weight that is used to control the impact of each particle's previous velocity on its current velocity, q_1 and q_2 are two positive acceleration coefficients; $\text{rand}(\bullet)$ and $\text{Rand}(\bullet)$ are two independent uniformly distributed random variables with range $[0, 1]$, g is the global best position for all particles in the swarm, and k represents the parameters of the neural network for the PSO to determine.

Finally, the new position of the i -th particle in the next generation can be determined by:

$$X_{(k)i}(t+1) = X_{(k)i}(t) + V_{(k)i}(t+1) \quad i = 1, 2, \dots, N \quad (10)$$

Notice that the value of each component in $V_{(k)i}$ should be limited to the range $[-v_{\max}, v_{\max}]$ to control possible excessive roaming of particles outside the search space. Each particle in the swarm continues to change its position in the search space based on its velocity calculated in (9) and (10) until the predefined termination criterion is satisfied. The termination criterion is usually the maximum number of generations.

2.3.2. Chaotic PSO Method

Liu et al. (2005) proposed a CPSO by combining a PSO with an adaptive inertia weight factor (AIWF) and a chaotic local search (CLS). The AIWF is used to encourage good particles to revise their exploration to refine their results through a local search and to encourage bad particles to exploit the search space with a large step. The AIWF is determined by:

$$l = \begin{cases} l_{\min} + \frac{(l_{\max} - l_{\min})(f_i - f_{\min})}{f_{\text{avg}} - f_{\min}}, & f_i \leq f_{\text{avg}} \\ l_{\max}, & f_i \geq f_{\text{avg}} \end{cases} \quad (11)$$

where l denotes the inertia weight that is used to control the impact of each particle's previous velocity on its current velocity, l_{\max} and l_{\min} are, respectively, the maximum and minimum of l , f_i is the current objective value of the i -th particle, and f_{avg} and f_{\min} are the average and minimum objective values of all particles, respectively.

A CLS is used to perform a locally oriented search (exploitation) for the global best solution f_i , which is the result of the PSO procedure. The CLS is based on the logistic

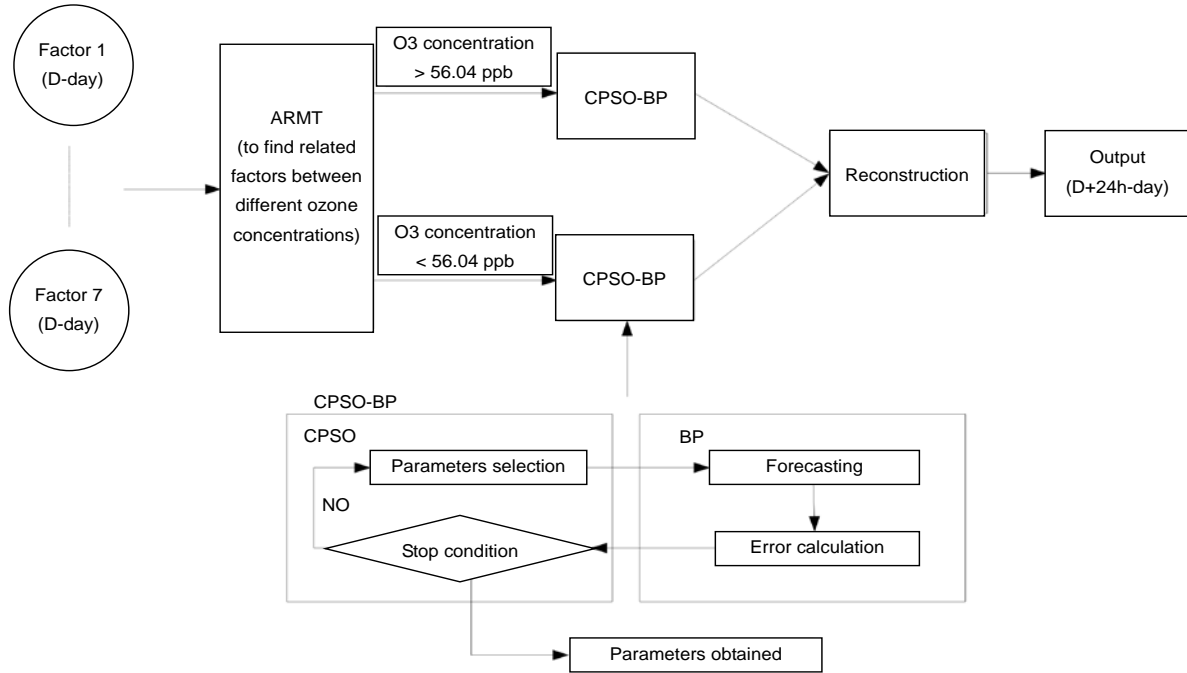


Figure 3. Structure of the ARMT-CPSO-BP neural network.

equation, with sensitive dependence on initial conditions and is defined as follows:

$$cx_{(k)i}^{(\lambda+1)} = 4cx_{(k)i}^{(\lambda)}(1 - cx_{(k)i}^{(\lambda)}) \quad i = 1, 2, \dots, N \quad (12)$$

where $cx_{(k)i}$ is the i -th chaotic variable, λ represents the iteration number, $cx_{(k)i}^{(\lambda)}$ is distributed in the range (0, 1), and $cx_{(k)i}^{(0)} \in (0, 1)$, but $cx_{(k)i}^{(0)} \notin (0.25, 0.5, 0.75)$ (Huang et al., 2008).

The improved PSO with chaotic local searching (CPSO) is employed to optimize the structure of the artificial network so that the point of minimum error can be found without adjusting the parameters of the neural network. A BP neural network is employed, and the CPSO is only applied to test its potential for determining the parameters of the BP neural network. Finally, the outputs of the CPSO neural network go through the reconstruction phase to generate final results.

2.4. ARMT-CPSO-BP Model

First, we use the association rule mining technique to find correlations between meteorological variables and ozone concentrations. Then, the significant correlation coefficients are fed into the hybrid CPSO-BP model to yield the predictions.

In this article, the proposed chaotic PSO is used to optimize the artificial neural network to avoid local extrema and to enhance the network's learning capability and training efficiency by defining the position vector function \vec{x} of the particle swarm as a network structure parameter and by improving the CPSO iterative process. As the network adopted in this paper evolves from the BP network, the weight w_{ij} from the hidden

layer to the output layer can be defined by the linear regression equation in the literature.

First, the position vector \vec{x} and the particle swarm are initialized, and the iteration process begins. Aimed at the network structure of each particle in each generation, learning is carried out to calculate the fitness value as defined by the following:

$$I = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^C (y_{j,i}^d - y_{j,i})^2 + \frac{1}{t_{mn}} \quad (13)$$

In the CPSO algorithm, the particle structure that minimizes the fitness function is the optimal network structure (Yang and Li, 2006).

In the above equation, N is the number of training samples, $y_{j,i}^d$ is the ideal output value of the i -th sample in the j -th layer output node, N is the number of network output neurons, and t_{mn} is the network structure training time for the n -th particle in the m -th generation. Because both learning capability and training efficiency, as indicated by different network structures, were considered when selecting the fitness function, the CPSO can optimize the neural network from all sides.

The ARMT-CPSO-BP model is applied to data taken from 1 July 2009 to 30 July 2009. The data from 1 July to 24 July are used as the training sample, and the data from 25 July to 30 July are used as the test sample. Using these sample data, define the particle \vec{x} in the particle swarm as the number of hidden neurons and the structure parameters. Set 30 as the number in the swarm. The initial parameters of the CPSO can be given as $V_i \in (-1.0, 1.0)$ and $C_1 = C_2 = 1.8$.

Set 5000 iteration steps as the condition to terminate the training of the network or 300 iteration steps for the particle swarm.

Then, we can use the ARMT-CPSO-BP method to predict ozone concentration. The structure of the ARMT-CPSO-BP model is shown in Figure 3.

2.5. Performance Indices

The mean bias error (MBE), which represents the mean of the bias of the predicted value with respect to the observed value. If O_i is the observed i -th value, P_i is the predicted i -th value, i.e., the estimate of O_i , and $i = 1, 2, \dots, N$, where N is the number of data samples, then MBE can be expressed as:

$$MBE = \frac{1}{N} \sum_{i=1}^N (P_i - O_i) \quad (14)$$

Note that a positive value for the MBE indicates an overprediction.

The mean absolute error (MAE), which represents the mean of the absolute bias of the predicted value with respect to the observed value. MAE can be expressed as:

$$MAE = \frac{1}{N} \sum_{i=1}^N |P_i - O_i| \quad (15)$$

The square root mean square bias error (RMSE), which represents the square root of the mean of the square of the bias of the predicted value with respect to the observed value. It can be expressed as:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (P_i - O_i)^2} \quad (16)$$

3. Case Study

3.1. Test Site

The data used for ozone concentrations in this study were obtained from the Semi-Arid Climate and Environment Observatory of Lanzhou University (SACOL), which is a participating site in the Coordinated Enhanced Observing Period (CEOP). CEOP was initiated as a major step towards bringing together the research activities of GHP/GEWEX and related projects by the WCRP (CLIVAR, CLIC) and has now become an important element of the WCRP. CEOP has been endorsed by the Integrated Global Observing Strategy Partnership (IGOS-P) as the first element of the IGOS-P Integrated Global Water Cycle Observations (IGWCO) theme.

SACOL is approximately 48 km from the center of Lanzhou, situated at the southern bank of the Yellow River in Gansu province in northwest China (35.57 °N, 104.08 °E). It is on China's Loess Plateau, approximately 1,965.8 m above sea level. The topography around the site is characterized by the Loess

Plateau, which consists of plains, ridges and mounds, etc., with elevations ranging from 1,714 to 2,089 m. The parent soil material is mainly quaternary Aeolian loess with the main soil type being sierozem. There is also some loess soil in the terraces, ridges and mounds. The terrain where the measurements are carried out is flat and covered with short grass (Huang et al., 2008).

3.2. Pollution around SACOL

The SACOL station has two main sources of pollution: a cement plant and a fertilizer plant. In addition, Lanzhou University as well as some local residents are located at the bottom of the hill, and there is a brick factory to the north. The school and the local residents contribute to the pollution, which is much worse in winter. Furthermore, the city of Lanzhou is about 48 km from the SACOL station, and there is an iron and steel plant about 3 km away. When there is a strong northwesterly wind, the pollutants from Lanzhou and the iron and steel plant can be swept to the SACOL site.

3.3. Instruments and Measurement

SACOL is equipped with ML9800 ambient air analyzers (Monitor Labs), which measure the concentrations of carbon monoxide (CO) (ML9830B), ozone (O₃) (ML9810B), sulfur dioxide (SO₂) (ML9850B), and nitrogen oxides (NO/NO₂/NO_x) (ML9841B) in the atmosphere.

3.4. Data Sets

The data sets contain both the concentration of air pollutants and the meteorological variables listed below:

- Surface wind speed
- Actual time temperature
- Relative humidity
- Nitrogen oxide concentration
- Carbon monoxide concentration
- Total radiation intensity
- Ozone concentration

Note that the forecasting system is built using data from 1 July 2009 to 24 July 2009. The data from 25 July 2009 to 30 July 2009 are employed to test the proposed model.

4. Results and Discussion

4.1. Independent Input Variables

None of the inference methods described above performs reliably if any of the input factors are missing from the model. On the other hand, if the model contains unnecessary factors, it will be unnecessarily complex and will predict less well than it otherwise could. We therefore briefly discuss the methods that help us develop an economical model. We screen to discover which factors contribute significantly to variations in the results.

Table 1. Meteorological Variable Assignments

Meteorological variable (unit)		1	2	3	4
Surface wind speed (m/s)	Label	1	2	3	4
	Range	< 1.69	1.69 - 2.59	2.59 - 3.94	> 3.94
Actual time temperature (°C)	Label	5	6	7	8
	Range	< 16.12	16.12 - 18.68	18.68 - 22.91	> 22.91
Relative humidity (%)	Label	9	10	11	12
	Range	< 33.92	33.92 - 56.33	56.33 - 72.16	> 72.16
Nitrogen oxide concentration (ppb)	Label	13	14	15	16
	Range	< 1.67	1.67 - 2.80	2.80 - 4.04	> 4.04
Carbon monoxide concentration (ppb)	Label	17	18	19	20
	Range	< 0.19	0.19 - 0.26	0.26 - 0.379	> 0.37
Total radiation intensity (w/m ²)	Label	21	22	23	24
	Range	< -1.36	-1.36 - 43.46	43.46 - 392.16	> 392.16
Ozone concentration (ppb)	Label	25	26	27	28
	Range	< 36.25	36.25 - 44.47	44.47 - 55.62	> 55.62
Ozone concentration_24h (ppb)	Label	29	30	31	32
	Range	< 35.62	35.62 - 44.62	44.62 - 56.04	> 56.04

The primary screening principle is that a variable should not be included in a model if it does not significantly increase the regression sum of squares due to regression. A careful and systematic approach asks whether a specific factor makes a significant contribution after accounting for other factors that may hide the importance of the factor within a group of factors. When the number of factors in a problem is small, it is usually possible to choose a suitable model. However, in problems with a large number of factors that are each potentially important for representing the conditional mean of the response variable, an automated procedure is needed.

We prepare seven factors for analysis. They are generally believed to be the primary factors that influence ozone concentration. The seven factors are surface wind speed (X_1), actual time temperature (X_2), relative humidity (X_3), nitrogen oxide concentration (X_4), carbon monoxide concentration (X_5), total radiation intensity (X_6), and the ozone concentration 24 hours in the future (X_7).

Second, to use association rule mining or pattern mining, preprocessing with normalization of data is required. For calculation, we assign numbers as labels for meteorological variables in different ranges. The association rule mining results are expressed as $x_1x_2 \dots \Rightarrow x_3x_4$. This means that when the x_1, x_2 , etc. factors appear, x_3, x_4 , etc. also appear. The meteorological variable assignments and the association rule mining results are shown in Tables 1 and 2, respectively.

The stepwise regression procedure combines forward selection with backward elimination. As forward selection progresses, factors selected early on may become redundant when related factors are selected during later steps. Therefore, in stepwise regression, backward elimination is performed after every forward selection step to remove redundant variables from the model. Forward regression and backward elimination steps are repeated until no further change can be made to the model.

The steps in stepwise regression are as follows:

Table 2. Association Rule Mining Results

	Association rule mining technique				Support	Confidence
1	8	13	32	\Rightarrow 9 24 28	0.0328	0.94444
2	9	13	24	\Rightarrow 8 28 32	0.0328	0.62963
3	9	13	32	\Rightarrow 8 24 28	0.0328	0.94444
4	13	28	32	\Rightarrow 8 9 24	0.0328	0.70833
5	13	24	28 32	\Rightarrow 8 9	0.0328	0.73913
6	9	13	28 32	\Rightarrow 8 24	0.0328	0.94444
7	9	13	24 32	\Rightarrow 8 28	0.0328	1
8	9	13	24 28	\Rightarrow 8 32	0.0328	0.62963
9	8	13	28 32	\Rightarrow 9 24	0.0328	0.94444
10	8	13	24 32	\Rightarrow 9 28	0.0328	1
11	8	9	13 32	\Rightarrow 24 28	0.0328	0.94444
12	8	9	13 24	\Rightarrow 28 32	0.0328	0.62963
13	9	13	24 28 32	\Rightarrow 8	0.0328	1
14	8	13	24 28 32	\Rightarrow 9	0.0328	1
15	8	9	13 28 32	\Rightarrow 24	0.0328	0.94444
16	8	9	13 24 32	\Rightarrow 28	0.0328	1
17	8	9	13 24 28 32	\Rightarrow 32	0.0328	0.62963

(1) Step 0 is defined as the initial model. The initial model always includes the regression intercept (unless the no-intercept option has been specified with respect to the real requirement). As for the backward-stepwise and backward-removal methods, the initial model also includes all the effects specified to be included in the analysis. The initial model for these methods is therefore the entire model. For the forward-stepwise and forward-entry methods, the initial model only includes the regression intercept (unless the no-intercept option has been specified). The initial model may also include 1 or more effects specified to be forced into the model. If j is the number of effects specified to be forced into the model, the first j effects specified to be included are entered into the model at Step 0. Any such effects are not eligible to be removed from the model during subsequent steps. Effects may also be specified to be forced

Table 4. Factors Associated with Different Ozone Concentrations

Ozone concentration	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
>56.04 ppb	temperature	Relative humidity	NO _x	radiation	O ₃ _24h ahead
<56.04 ppb	wind speed	CO	radiation		O ₃ _24h ahead

into the model when the backward-stepwise and backward-removal methods are used. As in the forward-stepwise and forward-entry methods, any such effects are not eligible to be removed from the model during subsequent steps.

(2) The forward-entry method is a simple model-building procedure. At each step after Step 0, the entry statistic is computed for each effect eligible for entry into the model. If no effect has an entry statistic value that exceeds the specified critical value for model entry, then stepping is terminated. Otherwise, the effect with the largest entry statistic value is entered into the model. Stepping is also terminated when the maximum number of steps is reached.

(3) The backward-removal method is also a simple model-building procedure. At each step after Step 0, the removal statistic is computed for each effect eligible for removal from the model. If no effect has a removal statistic value less than the critical value for removal from the model, then stepping is terminated. Otherwise, the effect with the smallest removal statistic value is removed from the model. Stepping is also terminated when the maximum number of steps is reached.

(4) The forward-stepwise method employs a combination of the procedures used in the forward-entry and backward-removal methods. At Step 1, the procedures for forward entry are performed. At any subsequent step where 2 or more effects have been selected for entry into the model, forward entry is performed if possible, and backward removal is performed if possible, until neither procedure can be performed, and stepping is terminated. Stepping is also terminated when the maximum number of steps is reached.

(5) The backward-stepwise method employs a combination of the procedures used in the forward-entry and backward-removal methods. At Step 1, the procedures for backward removal are performed. At any subsequent step where 2 or more effects have been selected for entry into the model, forward entry is performed if possible, and backward removal is performed if possible, until neither procedure can be performed, then stepping is terminated. Stepping is also terminated when the maximum number of steps is reached.

(6) Either critical *F* values or critical *p* values can be specified to control entry and removal of effects from the model. If *p* values are specified, the actual values used to control entry and removal of effects from the model are 1 minus the specified *p* values. The critical value for model entry must exceed the critical value for removal from the model. A maximum number of steps can also be specified. If not previously terminated, stepping stops when the specified maximum number of steps is reached (Hsieh and Lu, 2008).

The results obtained from the stepwise regression model are shown in Table 3. *R* is a multiple correlation coefficient.

In this case, the factors are surface wind speed, carbon monoxide concentration, total radiation intensity, and ozone concentration 24 hours in the future. Table 4 gives the factors associated with different ozone concentrations.

Table 3. Regression Model Results

Step	Set of variables	R ²	RMSE
0	X7	0.286	11.04
1	X7, X6	0.341	10.6
2	X7, X6, X1	0.355	10.5
3	X7, X6, X1, X5	0.364	10.4

Table 4 suggests that when ozone concentration is higher than 56.04 ppb, the factors affecting ozone concentration are temperature, relative humidity, nitrogen oxide concentration, total radiation intensity, and O₃_24h ahead. Similarly, when ozone concentration is lower than 56.04 ppbv, the factors affecting ozone concentration are wind speed, CO, radiation, and O₃_24h ahead. In our model, the input factors are different based on whether ozone concentration is higher or lower than 56.04 ppb.

The forecasting system was built using data from 1 July 2009 to 24 July 2009. The data from 25 July 2009 to 30 July 2009 is employed to test the proposed model.

4.2. Results and Discussion

To compare these models, we empirically examined their relative effectiveness by applying them to predict ozone concentrations using the test data from 1-24 July 2009. The comparison of the performance of the three models is shown in Figure 4. The performance indices for the three models are shown in Table 5.

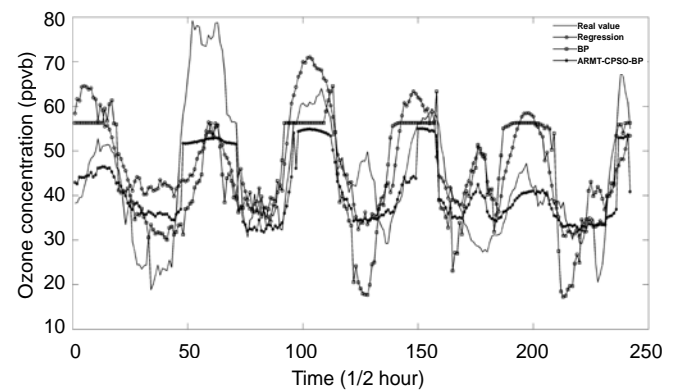


Figure 4. Comparison of the three models for the prediction of ozone levels using the test data set.

Forecasting error was analyzed for a total of 5 days, from 25 July 2009 to 30 July 2009. The results of each model can be compared for each site. A range of evaluative error statistics was used, as listed in Table 5. Mean estimated ozone concentrations from all methods do not differ substantially from the observed mean. The difference in mean, or mean bias error (MBE), is highest using the ARMT-CPSO-BP model, which under-predicts mean concentrations by 3.4 ppb. For the other models, the MBE is small. The observed ozone concentrations for the five days ranged from 18.8 to 79.2 ppb. For the similar BP model, the range is from 17.2 to 64.5 ppb. In three cases, the BP model performs better than the others. But as shown in Figure 4, the observed maximum and minimum values occurred at different times than the maximum and minimum values of the BP model. Therefore, these results do not indicate that the BP model is the best for forecasting ozone concentrations.

Table 5. Performance Indices of the Three Models

Model	Real	Regression	BP	ARMT-CPSO-BP
MBE	0.0	2.6	-0.08	-3.4
MAE	0.0	9.8	11.0	7.2
RMSE	0.0	12.2	13.6	9.3
MIN	18.8	30.8	17.2	31.8
MAX	79.2	71.0	64.5	55.0
RANGE	60.3	40.2	47.3	23.2
MEAN	44.4	46.9	44.2	40.9
Std. DEV.	13.4	10.0	11.5	7.2

The model results can be compared using measures of error or difference that are based on the deviations between predicted values and actual observations. The mean absolute error (MAE) is the average absolute value of these residual values, and the root mean square error (RMSE) is the square root of all squared residuals. MBE is easier to interpret and is sensitive to outliers, but the RMSE is more widely used and can be employed in further statistical analysis. The MAEs range from 7.2 ppb for the ARMT-CPSO-BP model to 11.0 ppb for the BP model. The ARMT-CPSO-BP model results in a slightly reduced MAE. The RMSE also indicates superior performance by the ARMT-CPSO-BP model. Each error measurement is proportional to the mean of each observed value. To identify which model has the least relative error, the RMSE or MAE can be expressed as a percentage of the mean, as shown in Table 5. In most cases, modest improvement is obtained using the ARMT-CPSO-BP model. This indicates that the ARMT-CPSO-BP model is better than the other models. The ARMT-CPSO-BP model achieves a 31.6% reduction in RMSE compared to the BP model and a 23.8% reduction compared to the regression model.

Ozone concentrations are usually highest at noon. The daily trend is usually a steady increase in ozone levels from the start of the day to 13:00, followed by a tapering through the afternoon and night. Seasonal peaks in ozone concentration occur at approximately 13:00. However, not all days have the same trend. Rain distribution at some sites can affect the trend. For example, if there is rain one day, the ozone level will not increase as much. However, we do not include rain in the training data, so the predictions are rarely accurate for these un-

sual samples. In the semi-arid areas of China, rain levels are always 0.00 mm, so including rain in the training data will not produce good results.

Because the data are for an urban context, there are unusual sources of ozone; thus, we can regard ozone as a disciplinary variable. Moreover, there is no guarantee of the model's reliability once it is extrapolated beyond the range of the input data used to construct it; that is, the model has only been validated for this time period and may not be appropriate if the meteorological conditions were unusual.

Another consideration in applying our model is the representativeness of the variable constructs. Ozone formation and variability result from photochemical reactions involving the interplay of many factors including meteorological, topographical, and vehicle emission variables. This study included eight vehicle emission and meteorological variables in considering various conditions. However, by using the association rule mining technique in the ARMT-CPSO-BP model, we avoid this problem to some degree.

Moreover, like all quantitative methodologies, the ARMT-CPSO-BP model requires an experienced or informed practitioner. The prediction models developed in this study employed a large number of training samples over 30 days. On each of these days, ozone concentration was measured 48 times; thus, the training samples have adequate data. The use of the forecast day's values does not seriously limit the usefulness of the model for predicting ozone.

5. Conclusions

This paper proposes a new hybrid forecasting method, namely, the ARMT-CPSO-BP neural network, for the prediction of ozone concentrations in semi-arid regions of China. The method is developed based on the association rule mining technique, a chaotic particle swarm optimization algorithm and a BP neural network. Performance evaluation indices are also introduced. The method is implemented along with a traditional regression model and a BP model. The results of the proposed method are compared with results obtained from the regression model and the BP neural network. The comparison shows that the proposed method produces satisfactory results in forecasting ozone levels and is superior to both the BP neural network and the regression model. The hybrid model achieves a 31.6% reduction in RMSE compared to the BP model and a 23.8% reduction compared to the regression model. The hybrid model is a promising candidate for forecasting rural ozone concentrations.

Acknowledgments. This research was supported by the National Basic Research Program of China '973' Program (Grant No. 2011CB706900), Scholarship Award for Excellent Doctoral granted by Lanzhou University, and Opening Fund of Key Laboratory of Land Surface Process and Climate Change in Cold and Arid Regions (Grant No. LPCC201201), CAS, China. Special thanks to the Semi-Arid Climate and Environment Observatory of Lanzhou University (SACOL) for providing us with the data. The authors also thank the anonymous reviewers and the editor for their valuable comments.

References

- Abdel-Aal, R.E., Elhadidy, M.A., and Shaahid, S.M. (2009). Modeling and forecasting the mean hourly wind speed time series using GMDH-based abductive networks. *Renewable Energy*, 34(7), 1686-1699. <http://dx.doi.org/10.1016/j.renene.2009.01.001>.
- Agrawal, R., Imielinski, T., and Swami, A. (1993). Mining association rules between sets of items in large database. Proceedings of the 1993 ACM SIGMOD international conference on Management of data, New York, 207-216. <http://dx.doi.org/10.1145/170035.170072>.
- Benton, J., Fuhrer, J., Gimeno, B.S., Skärby, L., Palmer-Brown, D., Ball, G., Roadknight, C., and Mills, G. (2000). An international cooperative programme indicates the widespread occurrence of ozone injury on crops. *Agric., Ecosyst. Environ.*, 78(1), 19-30. [http://dx.doi.org/10.1016/S0167-8809\(99\)00107-3](http://dx.doi.org/10.1016/S0167-8809(99)00107-3).
- Cai, J.J., Ma, X.Q., and Li, L.X. (2007). Chaotic particle swarm optimization for economic dispatch considering the generator constraints. *Energy Convers. Manage.*, 48(2), 645-653. <http://dx.doi.org/10.1016/j.enconman.2006.05.020>.
- Chen, J.L., Islam, S., and Biswas, P. (1998). Nonlinear dynamics of hourly ozone concentrations: Nonparametric short term prediction. *Atmos. Environ.*, 32(11), 1839-1848. [http://dx.doi.org/10.1016/S1352-2310\(97\)00399-3](http://dx.doi.org/10.1016/S1352-2310(97)00399-3).
- Clarke, B.B., Greenhalgh-Weidman, B., and Brennan, E.G. (1990). An assessment of the impact of ambient ozone on field-grown crops in New Jersey using the EDU method: Part 1 - white potato (*Solanum tuberosum*). *Environ. Pollut.*, 66(4), 351-360. [http://dx.doi.org/10.1016/0269-7491\(90\)90150-B](http://dx.doi.org/10.1016/0269-7491(90)90150-B).
- Dutot, A.L., Rynkiewicz, J., Steiner, F.E., and Rude, J. (2007). A 24-h forecast of ozone peaks and exceedance levels using neural classifiers and weather predictions. *Environ. Model. Software*, 22(9), 1261-1269. <http://dx.doi.org/10.1016/j.envsoft.2006.08.002>.
- Elkamel, A., Abdul-Wahab, S., Bouhamra, W., and Alper, E. (2001). Measurement and prediction of ozone levels around a heavily industrialized area: a neural network approach. *Adv. Environ. Res.*, 5(1), 47-59. [http://dx.doi.org/10.1016/S1093-0191\(00\)00042-3](http://dx.doi.org/10.1016/S1093-0191(00)00042-3).
- Feng, Z.Z., and Kobayashi, K. (2009). Assessing the impacts of current and future concentrations of surface ozone on crop yield with meta-analysis. *Atmos. Environ.*, 43(8), 1510-1519. <http://dx.doi.org/10.1016/j.atmosenv.2008.11.033>.
- Geethanjali, M., Mary Raja Slochanal, S., and Bhavani, R. (2008). PSO trained ANN-based differential protection scheme for power transformers. *Neurocomputing*, 71(4), 904-918. <http://dx.doi.org/10.1016/j.neucom.2007.02.014>.
- Hsieh, K.L., and Lu, Y.S. (2008). Model construction and parameter effect for TFT-LCD process based on yield analysis by using ANNs and stepwise regression. *Expert Syst. Appl.*, 717-724. <http://dx.doi.org/10.1016/j.eswa.2006.10.014>.
- Huang, J.P., Zhang, W., and Zuo, J.Q. (2008). An overview of the semi-arid climate and environment research observatory over the Loess Plateau. *Adv. Atmos. Sci.*, 25(6), 906-921. <http://dx.doi.org/10.1007/s00376-008-0906-7>.
- Jaffe, D., and Ray, J. (2007). Increase in surface ozone at rural sites in the western US. *Atmos. Environ.*, 41(26), 5452-5463. <http://dx.doi.org/10.1016/j.atmosenv.2007.02.34>.
- Kennedy, J., and Eberhart, R. (1995). Particle swarm optimization. Proceedings of the 4th IEEE International Conference on Neural Networks. Piscataway: IEEE Service Center. 1942-1948.
- Liu, B., Wang, L., Jin, Y.H., Tang, F., and Huang, D.X. (2005). Improved particle swarm optimization combined with chaos. *Solitons Fractals*, 25(5), 1261-1271. <http://dx.doi.org/10.1016/j.chaos.2004.11.095>.
- Lu, H.C., Hsieh, J.C., and Chang, T.S. (2006). Prediction of daily maximum ozone concentrations from meteorological conditions using a two-stage neural network. *Atmos. Res.*, 81(2), 124-139. <http://dx.doi.org/10.1016/j.atmosres.2005.11.007>.
- Martínez-Ballesteros, M., Martínez-Álvarez, F., Troncoso, A., and Riquelme, J.C. (2011). An evolutionary algorithm to discover quantitative association rules in multidimensional time series. *Soft Comput.*, 15(10), 2065-2084. <http://dx.doi.org/10.1007/s00500-011-0705-4>.
- Martínez-Ballesteros, M., Troncoso, A., Martínez-Álvarez, F., and Riquelme, J.C. (2010). Mining quantitative association rules based on evolutionary computation and its application to atmospheric pollution. *Integr. Comput. Aid. Eng.*, 17(3), 227-242. <http://dx.doi.org/10.3233/ICA-2010-0340>.
- Meleux, F., Solmon, F., and Giorgi, F. (2007). Increase in summer European ozone amounts due to climate change. *Atmos. Environ.*, 41(35), 7577-7587. <http://dx.doi.org/10.1016/j.atmosenv.2007.05.048>.
- Pérez, P., and Reyes, J. (2006). An integrated neural network model for PM10 forecasting. *Atmos. Environ.*, 40, 2845-2851. <http://dx.doi.org/10.1016/j.atmosenv.2006.01.010>.
- Prybutok, V.R., Yi, J.S., and Mitchell, D. (2000). Comparison of neural network models with ARIMA and regression models for prediction of Houston's daily maximum ozone concentrations. *Eur. J. Oper. Res.*, 122(1), 31-40. [http://dx.doi.org/10.1016/S0377-2217\(99\)00069-7](http://dx.doi.org/10.1016/S0377-2217(99)00069-7).
- Tang, X.L., Zhuang, L., Cai, J., and Li, C.B. (2010). Multi-fault classification based on support vector machine trained by chaos particle swarm optimization. *Knowledge-Based Syst.*, 23(5), 486-490. <http://dx.doi.org/10.1016/j.knsys.2010.01.004>.
- Tanimoto, H. (2009). Increase in springtime tropospheric ozone at a mountainous site in Japan for the period 1998-2006. *Atmos. Environ.*, 43(6), 1358-1363. <http://dx.doi.org/10.1016/j.atmosenv.2008.12.006>.
- Trevino, C., and Mendez, F. (1999). Simplified model for the prediction of ozone generation in polluted urban areas with continuous precursor species emissions. *Atmos. Environ.*, 33(7), 1103-1110. [http://dx.doi.org/10.1016/S1352-2310\(98\)00304-5](http://dx.doi.org/10.1016/S1352-2310(98)00304-5).
- Wang, Y.H. (2009). Nonlinear neural network forecasting model for stock index option price: Hybrid GJR-GARCH approach. *Expert Syst. Appl.*, 36(1), 564-570. <http://dx.doi.org/10.1016/j.eswa.2007.09.056>.
- Wilson, I.D., Paris, S.D., Ware, J.A., and Jenkins, D.H. (2002). Residential property price time series forecasting with neural networks. *Knowledge-Based Syst.*, 15(5), 335-341. [http://dx.doi.org/10.1016/S0950-7051\(01\)00169-1](http://dx.doi.org/10.1016/S0950-7051(01)00169-1).
- Yang, S.D., and Li, X. (2006). A New ANN optimized by improved PSO algorithm combined with chaos and its application in short-term load forecasting. *International Conference on Computational-Intelligence and Security*, Guangzhou, 945-948.
- Zhang, Y.D., and Lenan, W. (2009). Stock market prediction of S&P 500 via combination of improved BCO approach and BP neural network. *Expert Syst. Appl.*, 36(5), 8849-8854. <http://dx.doi.org/10.1016/j.eswa.2008.11.028>.